VisTrails: Using Provenance and Workflows for Scientific Exploration

David Koop
University of Massachusetts-Dartmouth

The mountains of data being gathered and generated each day have brought about many opportunities to discover and test new ideas. Algorithms help to summarize, filter, and mine this data, but insight often requires users to explore intermediate results and modify computations during their analyses. Visualization techniques provide intuitive and interactive methods for examining data, helping to speed the time to insight. Scientific workflows orchestrate and abstract complex data analysis pipelines, providing a structure that aids in understanding, reproducibility, and extensibility. While all of these ingredients help accelerate data analysis and exploration, keeping track of all of the computations and explorations places a burden on users. Automated methods that transparently capture provenance---the record of how each result was achieved---allow users greater freedom to investigate data without worrying about manually recording their progress. This talk will discuss how VisTrails supports scientific exploration by orchestrating tools and libraries, capturing provenance, and enabling reproducibility and reuse.